数理解析研究所講究録 1264 巻 2002 年 263-269

Why there isn't a complete description of the human society, I: The Individual and Rationality*

大阪大学大学院経済学研究科　　浦井　憲 (Ken URAI)

Graduate School of Economics, Osaka Univ.†

Abstract

In this paper, a formal set theoretical limitation in describing the society which consists of rational individuals is studied. It is shown that the concept of rationality, at least as a rational acceptability of sentences in a certain formal language, cannot be (introspectively) complete as long as we require it to be (logically) consistent. The result is obtained as a generalized version of Tarski's truth definition theorem, a closely related result to Gödel's second incompleteness theorem.

Keywords: Rationality, Recognition, Mathematical logic, Theory of sets, Game theory, Microeconomic theory, Tarski's truth definition theorem, Gödel's second incompleteness theorem.

JEL classification: A10; B40; C60; C70; D00

* Research was supported by the Japanese Ministry of Education Grant 12730013. I am indebted to professors Kiyoshi Kuga (Osaka University), Hiroaki Nagatani (Osaka University), and Jun Iritani (Kobe Univertisy) for helpful comments on several topics in this research program and, especially, for many discussions at the seminar on mathematical logic during 1995 – 1998. My special thanks are also due to professor Mamoru Kaneko (Tsukuba University) with his resent reseaches on the common knowledge and the foudation of game theory the author is much inspired. Preliminary versions of this paper were presented in a conference at the Research Institute for Mathematical Sciences, Kyoto University (December 9, 2001) and Osaka-Kobe joint seminar on mathematical economics at the Graduate School of Economics, Osaka University (December 22, 2001). The author also thanks to the participants at these seminars, my colleagues, and my laboratory members.

† Graduate School of Economics, Osaka University, Toyonaka, Osaka 560-0043, JAPAN. (Internet e-mail address: urai@econ.osaka-u.ac.jp)

1. INTRODUCTION

In the following, it will be shown that there is no set theoretical formal description of the human society that incorporates a quite natural and important kind of our inference (recognition) ability. There are many reasons for which we have to resign ourselves to obtain a complete economic model. Indeed, the standard economic theory admits that there are many types of 'externality'. Moreover, there are many unknown structures in the real world, especially in technologies, informations, preferences and expectations, etc. It seems, however, that such problems are treated by economic theorists as merely the gap between an idealized economic model and the reality. What I have concerned with here is not the gap between them but the impossibility of the notion of an *idealized model* itself.

If it is a purpose of economics to describe the human society as a theoretical and well founded mechanisms of 'rational' individuals. An economic model should formalize a system of rules which make each agent's behavior to be called 'rational'. In order to formalize such economic 'rationality', however, we should premise a restricted view on individual prospects or thoughts about the whole world. If it were not, as we shall see in this paper, the view of the world necessarily be inconsistent (hence, every action would be rational for him). On the other hand, with such a restricted view of the world, agents are not allowed to ask whether the world is exactly like what they are thinking about (in their view of the model). In other words, a consistent view (description) of the world should be incomplete in the sense that every agents should convinced in the rightness of the view itself without any proofs.

The reader may consider that the result in this paper is related to Gödel's second incompleteness theorem. Indeed, the main theorem in this paper may be considered as a generalized version of Tarski's truth definition theorem which is known as another important result of Gödel's lemma for the incompleteness theorem.[1] It should be noted, however, that there is an important difference between the foundation of mathematics and the foundation of our view on the society including ourselves. The former is the problem on what mathematics can do to formalize our rationality, and the latter is an argument of formalizing our rationality itself. We may change and reconstruct mathematics through our conviction and beliefs. In order to formalize ourselves, however, any restricted formalization may fail to characterize our total recognition ability; whereas there isn't any simple way or regular routine to formalize our general intelligence.

In this paper, the *rationality* is treated as an attitude to accept a formal assertion that is written in a certain formal language.[2] The syntax for such a language and semantics (especially for the meanings of the rationality) are given by a certain theory of sets $\mathscr{B} = (L_B, R_B, T_B)$ which we call an *underlying theory of sets*.[3] We assume that each person i using his/her formal language has a theory $\mathscr{L}_i = (L_i, R_i, T_i)$ which is at least as strong as the underlying theory of sets, $\mathscr{B}$.[4] Thus, we are modeling the situation that person i is possible to treat his/her assertion θ in the language of i (theory $\mathscr{L}_i$) as a set theoretic object $\ulcorner\theta\urcorner$ through the basic underlying theory $\mathscr{B}$. The problem we treat in this paper is that whether we may construct a *formula* $P_i(x)$ *of person* i in one free variable x such that $P_i(\ulcorner\theta\urcorner)$ means that θ is a rationally acceptable assertion of i. Of course the answer depends on properties that we request for the meanings of the 'rational acceptability'. What we have concerned with here are the logical *consistency*

[1] Mathematical concepts in this paper may be found in the standard literature in mathematical logic and/or theory of sets, e.g., see Kunen (1980), Jech (1997), Fraenkel, Bar-Hillel, and Levy (1973).

[2] Throughout this paper, we use such a linguistic definitions and approaches that may be common in standard arguments in the philosophical analysis. On the standpoint of our notions of rationality and truth, however, I depend much on the work of H. Putnum (1983).

[3] A precise definition will be given in section 2.

[4] Of course there must be an appropriate translation between his/her formal language and the language

($P_i(\ulcorner\theta\urcorner)$ and $P_i(\ulcorner\neg\theta\urcorner)$ never occurs simultaneously) and the introspective *completeness* ($P_i(\ulcorner\theta\urcorner)$ means $P_i(\ulcorner P_i(\ulcorner\theta\urcorner)\urcorner)$). The main theorem in this paper shows that there is no P_i satisfying both of these two important properties (Theorem 3). Theorems in this paper shows:

(1) The description of the world with the notion P_i cannot be a complete one as long as we need P_i to be consistent.

(2) Especially, we cannot introspectively recognize the consistency and the completeness (of our world view) itself.

(3) We cannot define (completely describe) what the rationality is as long as we require it to be consistent.

Therefore, all rational economic agent in a standard economic model, should believe in his/her rational choices without knowing what the rationality is. Every players in a non-cooperative game theory, should believe in his/her and other players' rational behaviors without knowing what the rationality exactly means. This seems to be a failure in all mathematical models of the social science based on the *methodological individualism*. Indeed, the concept of 'rational individual' (consistency) always prevent us from having a satisfactory answer to the question 'How the society is' (introspection) (see, Theorem 2, (b) and (c)). Of course this is not saying that all attempts in describing the society as the whole of rational individuals are meaningless. The result suggests, however, that such attempts never be completed even in an asymptotic sense and that we have to allow for the relation between our recognition abilities and views of the world.

2. THE WORLD VIEW

The difference between the approach in this paper and the ordinary economic model is that we request for an economic agent in the model to have a reasonable account for his/her economic behaviors. Let $I = \{1, 2, \ldots, m\}$ be the index set of agents. For each $i \in I$, denote by A_i the set of possible economic actions for agent i. Each action profile $(a_1, a_2, \ldots, a_m) \in \prod_{i\in I} A_i$ in the economy decides an economic consequence c_i in a set C_i for each $i \in I$.

A standard microeconomic theory and non-cooperative game theory start from such an individual decision making problem. In most economic models, there are stories or mathematical structures, e.g., equilibrium and solution concepts, that enable for each agent i to have a sufficient reason for his/her choices of an action a_i. As there are many reasons for (mutually exclusive) many actions to be chosen, there may also be many equilibrium and solution concepts. The rationality (the reason) in this sense crucially depends on the view of the world (the equilibrium concept). The purpose of this paper is to show that this type of rationality is completely different from our true rationality (thinking) and that the use (merely a part) of our true rationality may lead us to deny any such a specific view of the world and the rationality in the restricted sense.

In this paper, we suppose that agent i has a theory (written by a formal language) $\mathscr{L}_i = (L_i, R_i, T_i)$ for obtaining a reason to decide an action a_i. L_i is the list of all symbols for the language, R_i is the list of all syntactical rules including construction rules for terms, formulas, and all inference rules (making a consequent formula from original formulas, e.g., modus ponens, instantiation, etc.), and T_i is the list of all axiomatic formulas for the theory. We assume that each element of L_i may be uniquely identified with (coded into) an object in a certain basic theory of sets, $\mathscr{B} = (L_B, R_B, T_B)$, which we call an *underlying theory of sets* for $\mathscr{L}_i$.[5]

[5] Practically, the reader may identify $\mathscr{B}$ with Zermelo-Fraenkel set theory under the first order predicate logic. Since such a coding argument is usually restricted in the domain of finitistic objects, a minimal

The first important assumption of this paper is that such a set theory is so basic that every agent could develop (understand) it by their own language.

(A.1) The theory $\mathscr{L}_i = (L_i, R_i, T_i)$ is at least as strong as $\mathscr{B} = (L_B, R_B, T_B)$.[6] (Here, we implicitly assume that there is an appropriate translation between the languages for $\mathscr{L}_i$ and $\mathscr{B}$. Throughout this paper, such a translation is assumed to be fixed, and we suppose that each formula φ in $\mathscr{B}$ could be identified with "the same" formula in $\mathscr{L}_i$ without loss of generality.)

The second assumption in this paper is that though the theory, $\mathscr{L}_i = (L_i, R_i, T_i)$, of i may be stronger than $\mathscr{B} = (L_B, R_B, T_B)$, the structure of theory $\mathscr{L}_i$, i.e., each rules in list R_i is written in the language of the underlying theory of sets, $\mathscr{B}$. More precisely;

(A.2) $\mathscr{B}$ describes $\mathscr{L}_i$ in the following sense: (i) Each member of list L_i is a set in theory $\mathscr{B}$. (ii) List R_i consists of formulas in theory $\mathscr{B}$. Especially, there are two formulas in one free variable, $Term_i(x)$ and $Form_i(x)$, describing, respectively, the construction rules for terms and formulas of i. Every inference rule, as a relation among formulas of i, is also written in the language of $\mathscr{B}$. (iii) $Axiom_i(x)$ which defines formulas of i belonging to list T_i is a formula in $\mathscr{B}$.

Note that, under assumption (A.2), a combination of inference procedures, such as a proof procedure in theory $\mathscr{L}_i$, may be identified with a set theoretic procedure written in the form of a formula in theory $\mathscr{B}$. It should also be noted that each term, formula, and inference procedure (including the proof procedure) of i may not be finitistic (recursive) since the set theoretic methods in $\mathscr{B}$ may be much stronger than the finitistic method.

Under (A.1) and (A.2), an agent i is possible to treat an assertion (formula) θ in the language of i (theory $\mathscr{L}_i$) as a set theoretic object $\ulcorner\theta\urcorner$ through the underlying theory of sets, $\mathscr{B}$.[7] In the following, we call the theory, $\mathscr{L}_i = (L_i, R_i, T_i)$, satisfying these two assumptions, (A.1) and (A.2), the *world view* of i. The world view may include many features of the real world by adding additional axioms and syntactical rules, if necessary, and we suppose that an agent i chooses a 'rational' action $a_i \in A_i$ under the world view, $\mathscr{L}_i$. The third assumption is on the possibility of such a structure in the world view deciding the 'rationality'.

(A.3) There is a formula, $P_i(x)$, in one free variable, x, in the theory of i to mean that $x = \ulcorner\theta\urcorner$ for a certain formula θ of i and θ is *rationally acceptable* for i. The meaning of $P_i(x)$ as a way to decide such sentences is also given as a set theoretic property under the set theory $\mathscr{B}$, (hence, we may not require it to be finitistic), so that $P_i(x)$ may also be identified with a formula in $\mathscr{B}$.

Under (A.2), one of the most typical set theoretic procedure in $\mathscr{B}$ satisfying conditions in (A.3) for $P_i(x)$ (the rational acceptability) may be the proof procedure in $\mathscr{L}_i$, though we do not confine ourselves to this most familier case. In ordinary settings in economics, such a P_i may be considered as an arbitrary formula allowing, at least, one assertion specifying a certain character of $a_i \in A_i$ as a possible final decision of an agent i, as rationally acceptable. For example, such assertions may be: "final decision $a_i \in A_i$ of i is a price taking and utility maximizing behavior," for an ordinary micro economics settings, "final decision $a_i \in A_i$ of i is a best response given other agents' behaviors," for Nash equilibrium settings, and so on. It follows that, an agent $i \in I$ chooses an action $a_i \in A_i$ only if there is a sentence of i, θ, which is rationally acceptable, $(P_i(\ulcorner\theta\urcorner))$, asserting that agent i is allowed to chose action a_i as his/her final decision.

underlying set theory may be $ZF^- - P - INF$, ZF with the axiom of foundation, the power set, and the infinity are deleted.

[6] That is, every theorem in $\mathscr{B}$ is a theorem in $\mathscr{L}_i$.

[7] For finitistic objects, the notation $\ulcorner\ \urcorner$ is called Quine's corner convention.

3. THE RATIONALITY

As stated in the introduction, we are considering that an economic model should incorporate a structure which makes each agent's behavior to be called *rational*. In the previous section, such a structure is represented by the formula, $P_i(x)$, for agent i under the world view, $\mathcal{L}_i = (L_i, R_i, T_i)$, of i. We shall make in this section a further specification on the property $P_i(x)$, the *rationality* of i.

Perhaps, the most important property for P_i to be called as the rationality of i will be the consistency. It seems, however, that there are two kind of such consistency. One is the logical consistency and the other is the semantical consistency. We say that $P_i(x)$ is *logically consistent* if for any sentence θ of i, $P_i(\ulcorner\theta\urcorner)$ and $P_i(\ulcorner\neg\theta\urcorner)$ do not hold simultaneously. The logical consistency of $P_i(x)$ as a fact in the underlying theory of sets, $\mathcal{B}$, is denoted by $CONS(P_i)$. Formally;

(D.1) $CONS(P_i)$ is a formula in $\mathcal{B}$ which is equivalent to saying that $Form_i(\ulcorner\theta\urcorner) \rightarrow (P_i(\ulcorner\theta\urcorner) \rightarrow \neg P_i(\ulcorner\neg\theta\urcorner))$.[8]

The *semantical consistency* of P_i is the requirement that for any sentence θ of i, $P_i(\ulcorner\theta\urcorner)$ and $\neg P_i(\ulcorner P_i(\ulcorner\theta\urcorner)\urcorner)$ do not hold simultaneously. Since the condition (ordinarily) means that for each sentence θ of i, $P_i(\ulcorner\theta\urcorner) \rightarrow P_i(\ulcorner P_i(\ulcorner\theta\urcorner)\urcorner)$, we also call it the *introspective completeness* and denote it (as a fact in the underlying theory of sets) by $COMP(P_i)$. Formally;

(D.2) $COMP(P_i)$ is a formula in $\mathcal{B}$ which is equivalent to saying that $Form_i(\ulcorner\theta\urcorner) \rightarrow (P_i(\ulcorner\theta\urcorner) \rightarrow P_i(\ulcorner P_i(\ulcorner\theta\urcorner)\urcorner))$.

The logical consistency and the introspective completeness of P_i will be argued in the next section as mostly desirable properties for P_i. The reminder of this section is devoted to define additional basic properties for P_i. In the following, we assume that P_i automatically satisfies all of the following four properties.[9]

(A.4) If $\mathcal{B} \vdash \theta$, then $\mathcal{B} \vdash P_i(\ulcorner\theta\urcorner)$.

That is, each theorem in the underlying theory of sets is rationally acceptable for i.

(A.5) If $\mathcal{B} \vdash Form_i(\ulcorner\theta\urcorner) \wedge Form_i(\ulcorner\eta\urcorner) \wedge \ulcorner\theta\urcorner = \ulcorner\eta\urcorner$, then $\mathcal{B} \vdash P_i(\ulcorner\theta \leftrightarrow \eta\urcorner)$.

This implies that for each two formulas of i which are proved to be equal as set theoretical objects in $\mathcal{B}$, it is rationally acceptable to treat them as equivalent formulas.

(A.6) $\mathcal{B} \vdash Form_i(\ulcorner\theta\urcorner) \rightarrow (P_i(\ulcorner P_i(\ulcorner\theta\urcorner)\urcorner) \rightarrow P_i(\ulcorner\theta\urcorner))$.

The rational acceptability of θ under the rational acceptability of $P_i(\ulcorner\theta\urcorner)$ is quite natural.

(A.7) $\mathcal{B} \vdash (Form_i(\theta) \wedge Form_i(\eta)) \rightarrow (P_i(\ulcorner\theta \rightarrow \eta\urcorner) \rightarrow (P_i(\ulcorner\theta\urcorner) \rightarrow P_i(\ulcorner\eta\urcorner)))$.

If $\theta \rightarrow \eta$ and θ are rationally acceptable, then η is rationally acceptable. That is, the assumption means that rationally acceptable statements are closed under the modus ponens.

---

[8] Here, we implicitly assume that for each formula θ in $\mathcal{L}_i$, $\neg\theta$ is also a formula in $\mathcal{L}_i$, and that the translation process between $\ulcorner\theta\urcorner$ and $\ulcorner\neg\theta\urcorner$ may be written in a formula in $\mathcal{B}$. Note also that as stated in (A.3), $P_i(x)$ is considered as a formula in $\mathcal{B}$.

[9] The following assumptions are written in the form of theorems (or a metatheorems on theorems) in $\mathcal{B}$. The symbol $\vdash$ denotes that the right hand side is a theorem under the development of the theory that may uniquely be identified with the expression at the left hand side. Since proofs in $\mathcal{L}_i$ (hence, in $\mathcal{B}$,) may be considered as objects in the underlying theory of sets, an expression such as "$\mathcal{L}_i \vdash \theta$" may also be considered as a formula in the underlying set theory.

4. THE INCOMPLETENESS

In this section, the main result of this paper is given in the form of three theorems. These are different aspects of the same fact (a certain kind of incompleteness of P_i) under $\mathscr{B}$ with several auxiliary assumptions. The first theorem says that with additional properties in (A.1)–(A.7), $CONS(P_i) \wedge COMP(P_i)$ is false or is not rationally acceptable.

THEOREM 1. Under (A.1)–(A.7),[10]

$$\mathscr{B} \vdash (CONS(P_i) \wedge COMP(P_i)) \rightarrow \neg P_i(\ulcorner CONS(P_i) \wedge COMP(P_i)\urcorner),$$

PROOF. Let θ be a formula in one free variable in $\mathscr{L}_i$, $q(\ulcorner\theta\urcorner)$ be the formula $P_i(\ulcorner\neg\theta(\ulcorner\theta\urcorner)\urcorner)$, and Q be the formula $q(\ulcorner q\urcorner)$. Note that by (A.3), q, P_i, and Q may be considered as formulas in $\mathscr{B}$ as well as $\mathscr{L}_i$ though θ may not be. Moreover, for q to be well defined as a formula in $\mathscr{B}$, we assume (under condition (A.2)) that the procedure $\ulcorner\theta\urcorner \mapsto \ulcorner\neg\theta(\ulcorner\theta\urcorner)\urcorner$ may be written by the formula in $\mathscr{B}$. Then,

$$\mathscr{B} \vdash \ulcorner Q\urcorner = \ulcorner P_i(\ulcorner\neg Q\urcorner)\urcorner. \tag{1}$$

Since $\mathscr{B} \vdash (COMP(P_i) \wedge P_i(\ulcorner\neg Q\urcorner)) \rightarrow P_i(\ulcorner P_i(\ulcorner\neg Q\urcorner)\urcorner)$, by equation (1) together with (A.5) and (A.7), we have

$$\mathscr{B} \vdash (COMP(P_i) \wedge P_i(\ulcorner\neg Q\urcorner)) \rightarrow P_i(\ulcorner Q\urcorner). \tag{2}$$

Therefore,

$$\mathscr{B} \vdash (CONS(P_i) \wedge COMP(P_i)) \rightarrow \neg P_i(\ulcorner\neg Q\urcorner). \tag{3}$$

Then, by (A.4) and (A.7),

$$\mathscr{B} \vdash P_i(\ulcorner CONS(P_i) \wedge COMP(P_i)\urcorner) \rightarrow P_i(\ulcorner\neg P_i(\ulcorner\neg Q\urcorner)\urcorner). \tag{4}$$

As (1), it is also clear that $\mathscr{B} \vdash \ulcorner\neg Q\urcorner = \ulcorner\neg P(\ulcorner\neg Q\urcorner)\urcorner$. Then, by substituting it into (4),

$$\mathscr{B} \vdash P_i(\ulcorner CONS(P_i) \wedge COMP(P_i)\urcorner) \rightarrow P_i(\ulcorner\neg Q\urcorner). \tag{5}$$

Hence, by (3) and (5), we have

$$\mathscr{B} \vdash (CONS(P_i) \wedge COMP(P_i)) \rightarrow \neg P_i(\ulcorner CONS(P_i) \wedge COMP(P_i)\urcorner), \tag{6}$$

which was to be proved. ■

The next theorem consists of assertions with one more additional property, $CONS(P_i)$ or $COMP(P_i)$, to (A.1)–(A.7). The theorem shows how these two concepts are mutually introspectively inconsistent.

THEOREM 2. Assume that (A.1)–(A.7) hold.

(a) If $COMP(P_i)$, then $\mathscr{B} \vdash CONS(P_i) \rightarrow \neg P_i(\ulcorner CONS(P_i)\urcorner)$.
(b) If $CONS(P_i)$, then $\mathscr{B} \vdash COMP(P_i) \rightarrow \neg P_i(\ulcorner COMP(P_i)\urcorner)$.
(c) If $CONS(P_i)$, then $\mathscr{B} \vdash \neg COMP(P_i) \wedge P_i(\ulcorner\neg COMP(P_i)\urcorner) \wedge \neg P_i(\ulcorner COMP(P_i)\urcorner)$.
(d) If $COMP(P_i)$, then $\mathscr{B} \vdash \neg CONS(P_i) \wedge P_i(\ulcorner\neg CONS(P_i)\urcorner)$.

---

[10] More precisely, we are supposing that every facts in (A.1)–(A.7) may be treated as trivial theorems by definitions in the underlying theory of sets, $\mathscr{B}$.

PROOF. (a) and (b) may easily be obtained by deleting $COMP(P_i)$ (resp., $CONS(P_i)$) from (2) – (6) in the proof of Theorem 1. By (b), (A.4), and (A.7), we have

$$\mathscr{B} \vdash P_i(\ulcorner COMP(P_i)\urcorner) \rightarrow P_i(\ulcorner \neg P_i(COMP(P_i))\urcorner). \tag{7}$$

Moreover, by $CONS(P_i)$, we also have

$$\mathscr{B} \vdash P_i(\ulcorner \neg P_i(COMP(P_i))\urcorner) \rightarrow \neg P_i(\ulcorner P_i(\ulcorner COMP(P_i)\urcorner)\urcorner). \tag{8}$$

By (7) and (8), we have $\mathscr{B} \vdash P_i(\ulcorner COMP(P_i)\urcorner) \rightarrow \neg P_i(\ulcorner P_i(\ulcorner COMP(P_i)\urcorner)\urcorner)$, so that $\mathscr{B} \vdash \neg COMP(P_i)$. By using (A.4) and $CONS(P_i)$ repeatedly, we obtain (c). Lastly, by applying '(A.4) and (A.7)' twice on (a), we have

$$\mathscr{B} \vdash P_i(\ulcorner P_i(\ulcorner CONS(P_i)\urcorner)\urcorner) \rightarrow P_i(\ulcorner P_i(\ulcorner \neg P_i(\ulcorner CONS(P_i)\urcorner)\urcorner)\urcorner). \tag{9}$$

On the other hand, we have by (A.6)

$$P_i(\ulcorner P_i(\ulcorner \neg P_i(\ulcorner CONS(P_i)\urcorner)\urcorner)\urcorner) \rightarrow P_i(\ulcorner \neg P_i(\ulcorner CONS(P_i)\urcorner)\urcorner). \tag{10}$$

Therefore, by (9) and (10), we have $\mathscr{B} \vdash P_i(\ulcorner P_i(\ulcorner CONS(P_i)\urcorner)\urcorner) \rightarrow P_i(\ulcorner \neg P_i(\ulcorner CONS(P_i)\urcorner)\urcorner)$ so that $\mathscr{B} \vdash \neg CONS(P_i)$. Then, by (A.4), we have (d). ∎

The last theorem is on the inconsistency of all properties (A.1)–(A.7), $CONS(P_i)$, and $COMP(P_i)$, together with the underlying theory of sets, $\mathscr{B}$. It may also possible to understand the theorem as an undefinability theorem of the concept "rationality".

THEOREM 3. Under (A.1)–(A.7), $CONS(P_i)$, and $COMP(P_i)$, the theory $\mathscr{B}$ is contradictory.

PROOF. In this case, $\mathscr{B}$ proves $COMS(P_i) \wedge COMP(P_i)$ and $P_i(\ulcorner COMS(P_i) \wedge COMP(P_i)\urcorner)$ as well as Theorem 1. Hence, a contradiction follows. ∎

If we change (A.3) to assure the property of P_i in (A.3) without maintaining the existence of P_i, the above theorem asserts that there is no possibility for defining a concept of the rationality satisfying (A.4)–(A.7), $CONS(P_i)$ and $COMP(P_i)$, i.e., we obtain an *undefinability theorem of rationality*. The special case that $\mathscr{B} = \mathscr{L}_i = ZF$ and P_i is considered as a definition of "truth" (which clearly satisfies (A.4)–(A.7), $CONS(P_i)$ and $COMP(P_i)$) is Tarski's truth definition theorem (see, Kunen (1980), p.41).

(Graduate School of Economics, Osaka University, Japan)

References

Fraenkel, A. A., Y. Bar-Hillel, and A. Levy (1973) : *Foundations of Set Theory*, 2nd Revised Ed. (Elsevier/Amsterdam).

Jech, Thomas (1997) : *Set Theory*. 2nd ed. (Springer/Berlin).

Kunen, Kenneth (1980) : *Set Theory*: An Introduction to Independence Proofs, Studies in Logic and The Foundation of Mathematics, Vol. 102. (Elsevier/Amsterdam).

Putnum, Hilary (1983) : *Realism and Reason*, Philosophical Papers Volume 3. (Campridge University Press/New York).